

정보검색 시스템과 온톨로지

울산대학교 최호섭 · 옥철영*

1. 서론

인간의 사고방식을 컴퓨터가 이해할 수 있고, 이러한 인간과 컴퓨터의 의미적인 상호작용(semantic interaction)이 컴퓨터와 컴퓨터 사이에서도 그대로 반영될 수 있는 기술적 환경이 조성되도록 하는 것은, 인공지능 뿐만 아니라 정보검색, 지식기반 시스템 등이 궁극적으로 추구하는 연구 개발의 목적이라 할 수 있다. 특히 **수많은 지식 정보의 저장소로 생각할 수 있는 웹을 중심으로, 웹 상의 지식 정보에 인간의 사고방식을 반영할 수 있는 잘 정의된 의미(semantic)를 부여한다면, 현재 컴퓨터를 이용한 정보검색이나 문서 해석 및 이해 등이 지능화될 뿐만 아니라 나아가 인간의 사고방식과 같은 추론 기능까지 갖출 수 있는 시스템이 개발될 수 있을 것이다.**

정보검색 시스템은 사용자가 원하는 지식 정보를 얼마나 정확하고 빠르게 검색하여 의미있는 지식 정보를 제공할 수 있는가에 따라 시스템의 성능과 평가가 좌우된다고 할 수 있다. 즉, 정보검색 시스템은 사용자 측면에서는 다양한 정보검색을 통해 의미 있는 지식 정보를 제공받을 수 있도록 하는 매개체 역할을 담당하며, 개발자 측면에서는 정보검색에서의 질의 처리 성능, 대용량 정보 처리 방법, 색인 기술 등과 같이 지식 정보에 대한 표현, 저장, 조직, 접근 등의 기술적 처리의 효율성에 초점을 맞춘다. 최근 이러한 정보검색 시스템의 기술적·활용적 측면에서 강조되고 있는 것이 의미 있는 지식 정보를 어떻게 컴퓨터가 이해하고 처리할 수 있는가에 초점이 맞춰지고 있다.

이러한 연구 방향과 더불어 최근 국내 학계에서도 높은 관심을 보이면서 관련 분야에서 많은 연구가 진행 중인 것이 바로 Tim Berners-Lee가 1998년에 제안한 시맨틱 웹(semantic web)이다. 시맨틱 웹은 잘 알려진 바와 같이, 컴퓨터간의 정보교환을 가능하게 하며 웹 상의 데이터의 의미를 사람이 아닌 컴퓨터가 이해·처리할

수 있는 기술로서, 기존의 웹과 구분되는 것이 아니라 웹의 지식 정보에 잘 정의된 의미를 부여함으로써 웹 상의 수많은 데이터가 의미 있는 지식 정보로 표현될 수 있는 방법을 보여준다고 할 수 있다.

시맨틱 웹 기술의 등장은 정보에 대한 표현, 공유, 재사용 등에 인식을 증대시키면서 인공지능 분야에서 부분적으로 연구되었던 지식 표현(knowledge representation), 추론(inference), 의미망(semantic network), 온톨로지(ontology)에 대한 관심 영역을 이끌어 내는 계기가 되었다. 특히 온톨로지는 시맨틱 웹의 핵심 기술이자 핵심 구성 요소로서 자리를 잡아 가고 있으며, 국외에서는 특정 영역 중심의 domain 온톨로지를 비롯한 upper, core, task 온톨로지 등을 실제로 구축하여 시맨틱 웹 기술에 활용함과 동시에 여러 애플리케이션에의 활용 방법을 모색하고 있는 실정이다. 그러나 국내에서는 실질적인 온톨로지가 소규모 연구실에서의 실험적인 수준으로 구축되거나 이론적 수준에 그치고 있는 실정이다. 이러한 국내의 상황은 정보검색 시스템 관련 개발 기술에서 온톨로지를 어떻게 활용할 것인가에 대한 여러 가지 논의가 계속되고 실험적인 수준에 그치고 있는 것과 일맥상통한다.

본 고에서는 온톨로지에 대한 이론적 개념 정립과 더불어 온톨로지 구축 사례, 정보검색에서의 온톨로지 활용 동향 등을 간략하게 살펴보고자 한다.

2. 온톨로지 개념 정립

2.1 온톨로지에 대한 해석

'온톨로지'라는 용어에 대하여 Guarino, N.은 다음의 표 1과 같이 해석이 가능하다고 정리하고 있다[20].

위의 해석 중 가장 일반적으로 사용되는 것이 4 혹은 5의 해석으로, 대부분의 연구자들이 "an ontology is an explicit specification of a (shared) conceptualization"[25,26]라는 정의를 이용하고 있으며, 덧붙여 시맨틱 웹에서는 특정 목적과 영역의 중요성을

* 종신회원

강조하고 있다. 위의 일반적인 온톨로지의 정의를 세부적으로 살펴보면, 공유(share)라는 것은 개념이 해당 그룹 구성원 뿐만 아니라 컴퓨터 간에 합의된 지식에 바탕을 두고 있다는 것을 의미하고, 개념화(conceptualization)는 특정 목적을 위해서 표현하고자 하는 대상 세계에서 일어나는 현상에 연관된 개념들을 파악하기 위한 추상적 모델을 말한다. 또한 형식적(formal)이라는 것은 기계 가독형이어야 한다는 것을 의미하며, 명시적(explicit)이라는 것은 개념의 종류와 그들 간의 관계, 그리고 그 개념들의 사용에 있어서 주어지는 제한점들을 명백하게 정의하는 것을 의미한다[1,2].

표 1 Possible interpretations of the term "Ontology"

1. Ontology as a philosophical discipline
2. Ontology as an informal conceptual system
3. Ontology as a formal semantic account
4. Ontology as a specification of a "conceptualization"
5. Ontology as a representation of a conceptual system via a logical theory
 - 5.1 characterized by specific formal properties
 - 5.2 characterized only by its specific purpose
6. Ontology as a the vocabulary used by a logical theory
7. Ontology as a (meta-level) specification of a logical theory

이러한 온톨로지의 해석들을 통해 기존의 시소러스(thesaurus)나 어휘 데이터베이스(lexical database), 의미망 등과 같이 다양한 형태의 어휘 집합(lexical set)들을 온톨로지의 일종으로 판단하여 온톨로지를 구축하고 표현하는 경우도 있다. 즉 온톨로지가 가지고 있는 일반적인 구성 요소인 개념(concept), 관계(relation), 속성(property) 등이 표현되어 있는 어휘 집합(lexical set)들을 온톨로지화 또는 구축하는 사례를 살펴보면, 다음의 표 2와 같이 비슷한 형태이지만 다양하게 구축되고 있음을 알 수 있다[4].

표 2 Ontology Example

- | | |
|---------------------------------------|--|
| Ontology Example (in practice) | <ul style="list-style-type: none"> ● Simple concept hierarchies ● Semantic network ● Thesaurus ● Frame system ● Logical models ● Lexical field ● Category ● Taxonomies on the Web ● Catalogs for on-line shopping ● Domain-specific standard terminology ● Etc. |
|---------------------------------------|--|

이것은 시맨틱 웹에서의 웹 온톨로지(web ontology) 등장으로 인해, 기존의 시소러스나 의미망, 어휘 데이터베이스, 표준 분류 체계 등(예, WordNet, UMLS, UNSPSC, RosettaNet, etc.)을 RDFs, DAML+OIL, OWL과 같은 웹 온톨로지 언어를 이용하여 온톨로지화

표현하는 연구 개발이 활발하게 진행됨과 동시에, 기존의 계층적 구조로 표현된 어휘 집합들을 재사용하고 구축 시간을 단축하기 위한 구축 사례가 많아짐으로써 이러한 온톨로지의 실질 구축 사례가 다양한 형태로 표현되었다고 할 수 있다. 물론 표 2와 같은 사례들을 통해 온톨로지는 클래스(class)와 서브클래스(subclass)의 관계, 즉 개념간의 계층적 구조를 형성하고 있다는 점과 어떠한 의미적 관계(semantic relation)를 가지고 있다는 점에서는 위의 사례들이 공통적일 수 있으며, 웹 온톨로지 언어의 사용 여부와 활용적인 측면에서는 다르다고 할 수 있는 것이다.

2.2 온톨로지의 개념

2.1절과 같이 온톨로지에 대한 해석들과 형태적인 사례를 통해 온톨로지가 어떠한 형식으로 표현되고 있는지를 간략하게 살펴보았다. 그렇다면 온톨로지에 대한 해석만 가지고는 온톨로지를 파악하기가 쉽지 않을 뿐만 아니라, 실질적으로 온톨로지를 구축하는 측면에서도 기존의 시소러스 구축 원리나 표준 분류 체계 등의 구축 원리를 이용한다면 문제가 될 수 있다. 이러한 문제가 발생하는 것은 '개념'과 '관계', 그리고 '개념화'에 대한 이해 때문이라 할 수 있다. '개념'은 사전적으로 여러 관념 속에서 공통된 요소를 뽑아내어 종합하여 얻은 하나의 관념으로 정의될 수 있으며, '개념'은 일반적으로 언어로 표현된다. '관계'는 어휘의미론(Lexical Semantics)에서의 의미 관계(semantic relation)와 조금 폭넓은 의미에서의 개념 관계(conceptual relation)가 있다. 특히 시소러스나 WordNet에서 일반적으로 사용 중인 의미 관계는 상하 관계, 동의/유의 관계, 부분전체 관계 등이 보편적으로 사용되며, 개념 관계는 조금 넓은 의미에서 구성된 관계(has_member), 위치 관계(has_location) 등으로 세분화하여 개념간의 관계를 설정할 수 있다. '개념화'는 사물이나 표상을 어떤 성질·공통성·본질에 착안하여 그것을 추출하여 파악하는 과정으로서, 그 형태적 표현은 분류적 구조(classified structure)와 계층적 구조(hierarchical structure)로 나타날 수 있다[4].

이러한 개념화 양상은 기존의 시소러스와 의미망 등과 같은 계층적 어휘 집합과 온톨로지를 구분하기 어려운 것 중의 하나이다. 이러한 온톨로지에 대한 어느 정도의 개념 정립을 위하여, 그림 1을 통해 간략하게나마 시소러스와 온톨로지의 차이점을 확인함과 동시에 어떠한 연관성이 있는지를 확인할 수 있다[13]. 덧붙여 그림 2를 통해서도 커뮤니케이션을 위한 기본적인 온톨로지의 원리를 파악할 수 있다[9]. 이를 통해 온톨로지의 기본 개념을 어느 정도 이해할 수 있으리라 판단된다.

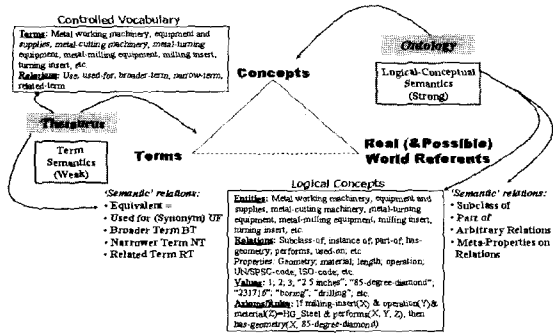


그림 1 Thesaurus versus Ontology

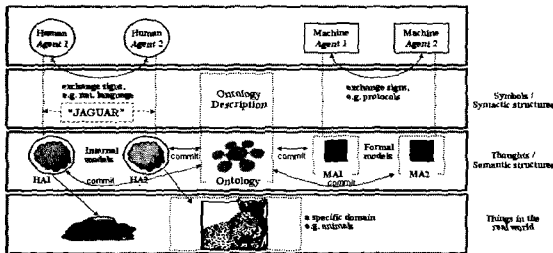


그림 2 Ontology for Communication

즉 그림 1, 그림 2를 통해 온톨로지가 단순히 용어 (term)의 체계적 구조화가 아니라 개념들을 특정한 영역의 개념에 대한 정의와 관계, 그리고 개념이 가지는 특수한 속성들로 이루어진 집합체임을 알 수 있다. 또한 온톨로지는 사람과 사람 사이의 원활한 커뮤니케이션처럼 기계와 기계 사이에도 형식적 모델을 통해 원활하게 커뮤니케이션이 가능하도록 하는 의미적인 구조를 가져야 하는 것도 알 수 있다.

이러한 온톨로지의 일반적인 개념 속에 추가적으로 고려해야 하는 것이 바로 추론과 통합(integration) 기능이다[9,10,11,31,32]. 온톨로지가 단순하게 계층적 구조와 의미 관계 등으로 구성되어 있다면, 온톨로지의 유용성은 시소러스와 의미망, 분류 체계, 어휘 데이터베이스 등과 다를 바 없을 것이다. 즉 온톨로지를 통해, 사람이 가지는 사고방식과 비슷한 추론 방법으로 새로운 개념을 유추하고 관계를 설정할 수 있어야 하는 것이다. 또한 다양한 온톨로지를 어떻게 통합할 수 있는가도 중요한 문제이다. 온톨로지의 정의에서도 알 수 있듯이 온톨로지의 특징 중에 하나인 공유는 온톨로지를 구축함에 있어 반드시 고려해야 하는 문제임으로, 여러 분야에서 다양하게 만들어진 온톨로지를 어떻게 통합하느냐에 따라 온톨로지의 통용성을 확인시킬 수 있을 것이다.

이상의 내용을 2.1절과 더불어 전산학적인 입장에서 정리한다면, 온톨로지를 다음과 같이 정리할 수 있을 것이다.

- 일정한 체계에 의한 어휘 사전이나 용어의 확보
- 특정 영역(specified domain) 뿐만 아니라 보편 영역(generic domain)의 기본 개념에 대한 정의

- 와 그들 간의 관계에 대한 명세화
- 개념, 관계, 속성들의 유기적인 집합
- 전산적 처리가 가능한 구조화와 구체화
- 공유와 재사용의 가능
- 논리적 추론
- 통합

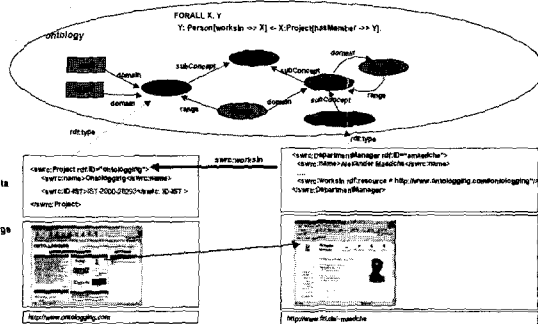


그림 3 Ontologies on the Web

2.3 온톨로지 구축 고려 사항

온톨로지의 구축 단계를 간략하게 설명하면, 특정한 목적과 영역을 고려한 다음, 개념을 자동·반자동 추출하거나 어휘 사전을 확보하여, 확보된 개념들을 정의하고 조직화해야 한다. 조직화한다는 것은 개념들 간의 관계를 설정함과 동시에 개념이 가지는 특수한 속성을 추출하여 체계화시키는 것을 의미하는 것으로서, 이론적 체계와 더불어 실질적인 구축 원리를 마련해야 한다. 다음 단계로 온톨로지를 표현할 웹 온톨로지 언어나 기계 가독형 표현 언어(machine readable representation language)를 설정하여 형식화하고 실질적으로 구현하여야 하며, 구축 중인 온톨로지와의 다른 온톨로지와의 통합 문제와 기존에 존재하는 많은 자원(resources)을 어떻게 활용할 것인가를 모색해야 한다. 그리고 마지막으로 구축된 온톨로지를 대상 애플리케이션(target application)에서 실험하거나 사용 패턴(usage patterns)을 분석하여 평가해야 한다. 또한 이러한 평가 결과를 바탕으로 유지·보수를 해야 하며, 조금 더 발전적인 온톨로지로서 개선해야 한다[4,10,15,27,28].

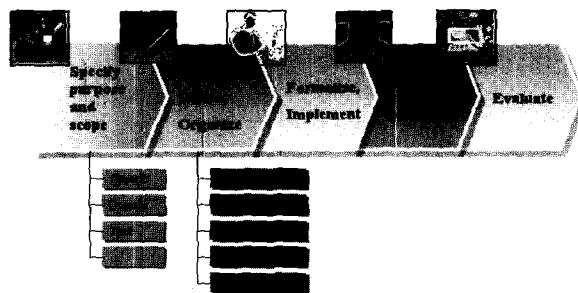


그림 4 General Ontology Development methodology

이러한 일련의 구축 단계에서 가장 어려운 문제는 바로 온톨로지를 실질적으로 구축하는 이론적 체계와 원리가 아직까지 마련되지 않았다는 것이다. 기존의 구축 사례들을 살펴보면, 온톨로지의 실질적인 구축 측면보다도 기존의 시소러스나 의미망, 분류 체계 등을 이용한 온톨로지 구축이나 기구축된 온톨로지를 이용한 애플리케이션 개발이 대부분을 차지한다. 이것은 온톨로지 구축에 있어서 국내외적으로 가지는 공통적인 문제라 할 수 있다. 이를 위해서는 WordNet, UMLS와 같이 관련 학문에 대한 이론 습득과 더불어 자연언어처리 기법의 활용을 통한 언어 습득 및 이해 처리 등과 같은 부수적인 연구가 뒤따라야 할 것이다.

이외에 온톨로지를 구축에 고려해야 될 사항을 정리하면 다음과 같다.

- 온톨로지 구축 대상자 선정
- 온톨로지 구축 방법(자동/반자동/수동)
- 응용분야별 온톨로지 구축 원리 파악
- 기구축된 시소러스, 분류 체계 활용 방법
- 범용적/국제적 수준의 온톨로지 구축
- 온톨로지 저작 도구 개발
- 온톨로지 구축 관련 표준 방식 연구
- 시맨틱 마크업 언어 표준안 개발
- 기타

3. 온톨로지 기반 정보검색 시스템

3.1 온톨로지 개발 및 활용 동향

시맨틱 웹의 등장은 온톨로지 자체에 대한 연구 개발 뿐만 아니라 온톨로지에 기반한 많은 애플리케이션 개발에도 영향을 미치기 시작했다. 이러한 온톨로지 연구 개발과 활용에 대한 대표적인 사례가 유럽을 중심으로 2000년부터 시작된 OntoWeb Project[29,34], Onto-Knowledge Project[32], 그리고 미국을 중심으로 한 W3C Semantic Web Activity(Web Ontology WG, Semantic Web Best Practices and Development WG...)[30] 등이라 할 수 있다.

'OntoWeb Project'는 프로젝트 명칭이 "지식 관리와 전자 상거래를 위한 온톨로지 기반 정보 교환(ontology-based information exchange for knowledge management and electronic commerce)"으로서, 온톨로지 기반 기술을 유럽 공동 시장에 제공할 수 있도록 함과 동시에 ISO, ANSI, W3C, IEEE 등에서 제시하는 국제 표준에 발맞추어 나가면서 이들 단체에게 표준화와 관련된 여러 정보를 제공할 수 있도록 여러 가지 온톨로지 기반 정보 교환 방법을 연구하는 장기적인 프로젝트

이다. 이 프로젝트에서는 다양한 온톨로지 기반 애플리케이션 개발 뿐만 아니라 평가 방법에 대한 연구를 비롯하여, 성공적인 시나리오와 가이드라인을 제공함과 동시에, 온톨로지 저작 도구(OntoEdit)[30] 등을 사용자에게 제공하고 있다. OntoWeb에서 온톨로지 활용하는 분야는 정보검색, 전자 상거래, 지식관리 등을 많은 응용 분야에서 테스트와 평가를 거치고 있다. 그림 5에서 OntoWeb 프로젝트의 개괄적인 내용을 확인할 수 있다.

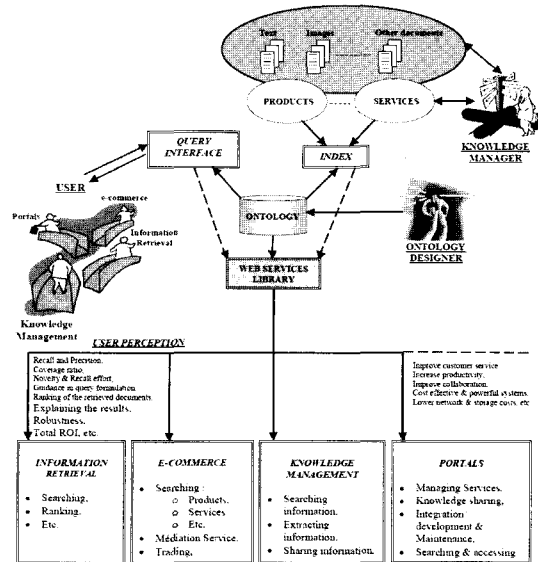


그림 5 Clusters of ontology-based application with some additional evaluation criteria(OntoWeb Project)

'On-To-Knowledge(OTK) Project'는 온톨로지를 이용한 내용 기반 지식 관리를 위한 각종 도구와 방법 등을 개발하기 위한 프로젝트로서, 특히 기업에서의 시맨틱 웹 기술 활용을 통한 지식 관리에 중점을 둔 연구 개발 프로젝트라 할 수 있다. 또한 OTK의 결과물을 살펴보면 시맨틱 웹 기반의 지식 관리 도구를 Intranet/WWW에서의 지식 정보 추출, 의미적 표현과 분석, 사용자 질의 제어 등을 고려한 도구를 실질적으로 개발하였다는 점에서 의의가 있다고 할 수 있다. OTK의 결과물 중 시맨틱 웹 기반의 지식 관리를 위한 도구는 그림 6에서 확인할 수 있으며, 그 외 OTK 프로젝트 연구자가 참여하여 개발된 웹 상에서 XML/RDF 기반 온톨로지와 추론 기능을 고려한 웹 온톨로지 언어인 OIL(Ontology Inference Layer), 지식 공학과 관리를 위한 방법론, 산업적 활용 가능성에 평가와 피드백 등이 프로젝트의 결과물로 제시되었다고 할 수 있다. 그리하여 OTK 프로젝트의 파트너였던 British Telecom Call Center, Swiss Life 생명보험회사 등에서 각각 Intranet 기반 가상 Community에서 온톨로지를 이용하

여 정보를 공유하거나, 대용량 문서에서 관련 정보를 검색하는 업무에 온톨로지를 이용함으로써 실제 기업에서의 활용과 평가도 확인할 수 있다.

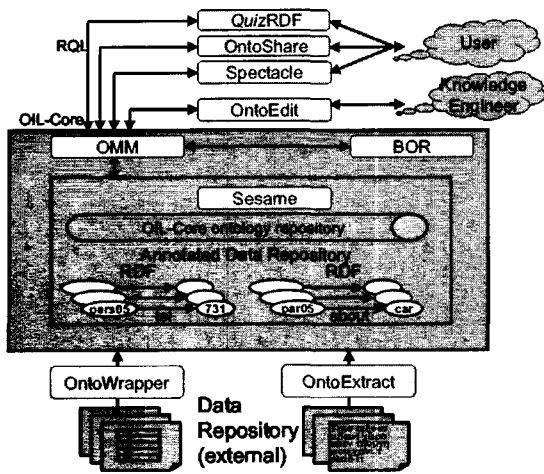


그림 6 Tool for Semantic Web-based knowledge management

'W3C Semantic Web Activity'는 W3C의 6개 분야 중 하나인 기술과 사회 분야(Technology and Society Domain)에서 다루는 웹 기술과 응용 표준 개발에서 필요한 6개의 활동 중 시맨틱 웹을 다루는 활동이다. 이 활동에는 세부적으로 Semantic Web Coordination Group, RDFCore Working Group(2001), RDF Data Access Working Group(2004), Web-Ontology Working Group(2001), Semantic Web Best Practices and Development Working Group(2004), Semantic Web Interest Group 등으로 형성되어 있다. 이 활동은 시맨틱 웹과 관련된 표준화와 더불어, 시맨틱 웹의 온톨로지 표준화에 앞장서고 있다. 특히 2002년 웹 온톨로지 언어인 OWL(Ontology Web Language)의 필요사항에 대한 초안 버전 1.0을 발표한 이후, 현재 많은 온톨로지 응용 분야에서 온톨로지를 OWL로 표현하고 있다.

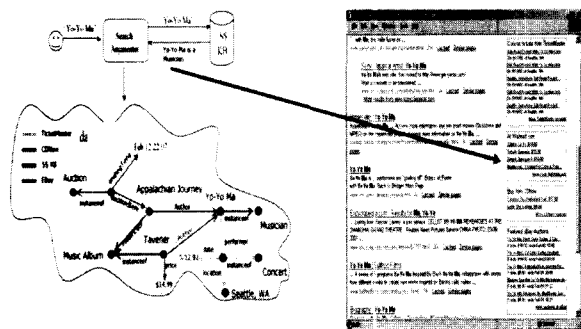


그림 7 Semantic Web Application: Semantic Search <TAP(building the semantic web) Project>[36]

3.2 정보검색과 온톨로지

최근 정보검색 기술은 랭킹 시스템, 중복 검색 결과 제거, 메타 검색, 분산·통합 검색, 전문 검색 등 다양한 분야의 정보들 중 사용자에게 더욱더 빠르고 정확하게 의미 있는 정보를 전달하고 하는 기술로 계속 발전하고 있다. 특히 최근에는 자연언어처리 기술을 이용한 의미 기반 정보검색 기술이나 질의응답 시스템 등이 활발하게 연구되고 있다[3].

이러한 연구 개발 흐름은 기존의 정보검색 기술 중의 하나였던 디렉토리 서비스, 키워드 기반 검색 등은 문서의 의미를 판단할 수 있는 기술은 아니었으므로, 그만큼 사용자의 의도에 맞는 정답 문서를 제공하지는 못하였다. 그리하여 사용자의 질의를 확장하기 위한 색인어 확장용으로 시소러스를 이용하여 사용자에게 의미 있는 문서를 제공하거나, 자연어 질의를 형태소 분석, 구문 분석, 의미 분석 단계를 거쳐 사용자의 의도를 분석하여 지식베이스와 추론 기법을 이용하여 의미 있는 문서를 제공하는 정보검색 기술이 이용되고 있다. 그러나 이러한 정보검색 기술 또한 아직까지 사용자의 질의 분석에서의 오류, 문서에서의 정확한 의미 분석 처리 문제, 단어 의미 중의성(word sense ambiguity) 문제 등 여러 가지 문제 때문에 실생활에서 사용하기에는 더 많은 시간이 필요할 것으로 보인다.

이러한 정보검색 기술 연구는 시맨틱 웹에서의 온톨로지 역할이 증가함에 따라, 현재의 웹 구조에 시맨틱 웹 기술을 결합한 시맨틱 웹 기반의 검색 시스템이 국내외적으로 개발 중에 있다. 즉 온톨로지 기반 정보검색 기술은 중요한 정보가 있는 자원을 빠르게 찾아 사용할 수 있다는 점과 자원을 찾는 정확도를 향상시킬 수 있다는 점에서 중요한 기술로 자리 잡아 가고 있다. 또한 검색 엔진이 온톨로지에 정의된 개념과 규칙을 활용하면, 온톨로지를 검색 향상을 위해 추론 규칙을 이용하기 때문에, 단순히 사용자의 질의와 일치되는 문서만 보여주는 것이 아니라 사용자의 질의의 의미를 분석하여 그와 관련된 정보를 온톨로지에 표현된 관계에 따라 다시 질의를 적절하게 바꿀 수도 있게 한다[2,14].

그러나 아직까지 국내외적으로 온톨로지 기반 정보검색 시스템이 아직 개발 단계에 있거나 시범적인 운용 서비스 단계에 있기 때문에 온톨로지를 이용한 실질적인 활용 실태를 정확하게 확인하기는 힘들지만 몇몇 프로젝트의 결과물이나 보고서, 시범 운용 서비스 사이트를 통해 그 내용을 확인할 수 있다.

온톨로지 기반 정보검색 시스템을 개발하기 위해서는 일련의 기술 정보를 마련해야 하는데, 그림 9는 'OntoWeb' 프로젝트에서 온톨로지 기반 정보검색 애플리케이션 개

발과 활용을 위한 가이드라인 항목과 정보검색 시스템 설계 단계에 필요한 주요 사항을 나타낸 것이다.

#	Guideline Item	Issues
1	Decide on the nature of the application	Will the ontology be used for • Descriptions? • Integration of heterogeneous info systems? • Better information retrieval? • Support of problem solving? Based on the intended use, a different set of evaluation criteria should apply.
2	Identify users and involve them from start	Because ontologies are about definitions and "understanding" between communicating parties, where humans are to be involved they constitute a fundamental issue. To improve transparency and hence acceptance of the system, its intended users must be consulted from the very beginning.
3	Look at company information/knowledge silos	Often groups within large organizations exist in "silos" where communication between them is non-existent or not very effective. If this is the case, are there duplicate databases, ontologies and how can the ontology application help bridge the gaps between the silos?
4	Decide on proprietary and non-proprietary concepts (parts of ontology)	If the intended application is of strategic importance, are there some parts of the ontology that must be kept company confidential? If this is the case, is a partitioning between the proprietary and public parts of the ontology possible? Can the non-proprietary part be found in the public domain and what are the implications in terms of integration with the proprietary part?
5	Examine your legacy systems	If the ontology application is not custom built check that integration with legacy systems is possible. This should go beyond the communications layer and consider the logical layers of the legacy and ontology systems.
6	Decide on Reasoning Processes	What kinds of use will the ontology be put to? If it to be used for reasoning and problem solving, are there any reasoning engines available or would it be necessary to develop one in-house? What is the state of the art and what the level of maturity of any existing technologies?
7	Define evaluation criteria and metrics	Although success will be most probably proved in practice, it is important to be able to quantify the advantages of the new system and hence its ROI. This should include the total effort invested and required for developing and maintaining the new solution.
8	Decide whether to cover all search modes	Information retrieval is not always performed with the same goals in mind and the range of expected useful answers might vary considerably for the same query. Each mode requires different indexing and scoring functions that should be explicitly encoded in the system from the beginning.
9	Consider how to deal with data noise	There are many applications that must assume noisy data (e.g. fault repair records, call center records etc.) with non-standard grammar and syntax (e.g. short hand). This noise could severely affect system performance and therefore should also be accounted for from the beginning of the process.
10	Research Existing Ontologies	Does an existing ontology exist that covers your application needs? This possibility should be considered seriously if the application will need to make use of industry wide information norms and if the information is not considered of strategic value to the company.
11	Look at Ontology Maintenance tools and procedures	If a specific ontology environment (that accompanies the ontology) is to be used examine the capabilities of the supporting tools, the assumptions of the underlying language and its expressive power. How easy is it to change concept definitions? Does the environment check for inconsistencies? Can it cope with multiple versions? Etc.
12	If you are to use external ontologies, look at what institutions / bodies share	Is the external ontology you have decided to use approved / backed by any industry wide body? Is there a maintenance procedure? Is there a working group that maintains the ontologies and allows users to voice their needs? Are other industry participants using this ontology?
13	Consider integration with your supplier / customer application	Is there any benefit in integrating the ontology application with that of your customers or suppliers? If so, do they have any equivalent applications in place? Is integration with your ontology possible and at what level?

그림 8 Checklist for the implementation of an ontology application

3.3 온톨로지 기반 정보검색 시스템

온톨로지 기반 정보검색 시스템을 개발하기 위해서는 그림 8과 같이 일련의 가이드라인을 기술해야 한다. 즉 온톨로지를 이용한 애플리케이션의 성질을 결정하고, 사용자와 관련된 사항을 확인하고, 특정 조직의 정보과 지식에 대한 분석, 소유권 문제, 특정 조직 내 시스템 조사, 추론 처리 결정, 평가 기준과 측정 기준 결정, 활용 가능한 범위 결정, Data noise 고려, 존재하는 온톨로지들에 대한 연구, 온톨로지 관리 도구와 처리 절차 분

석 등 여러 가지 사항들을 고려해야 한다.

온톨로지 기반 정보검색 시스템 개발을 위해서는 여러 가지 요소 기술들이 필요함과 동시에 실질적인 온톨로지 구축과, 구축 기관과의 연계, 기존 정보검색 기술과의 융합 기술, 개발 결과의 적극적 활용 방법 등 많은 부분들을 고려하지 않으면 안 된다. 이는 시맨틱 웹에서의 온톨로지 활용 방법이 주요 관심사가 되면서 이러한 온톨로지 기반 정보검색 시스템 개발은 앞으로 여러 연구 기관에서 대규모 또는 중소규모 실험 및 구축 개발이 진행될 것이다.

현재까지 연구 개발이 진행되고 있는 몇몇 온톨로지 기반 정보검색시스템의 사례를 통해, 실질적으로 온톨로지를 어떻게 정보검색에서 활용하고 있는지를 살펴보고 록 한다.

그림 9는 시맨틱 웹 기반 검색 시스템 구조를 설명한 것이다(1). 이 시스템 구조는 서브 시스템인 검색 엔진과 온톨로지 저장소(ontology repository)를 중심으로 상위가 검색 엔진이며, 하위가 온톨로지 시스템 구조이다. 검색 엔진은 사용자가 시맨틱 웹의 사용과 에이전트와 RDF, 온톨로지를 이용한 추론 엔진의 연동에 대한 서브시스템 구조이며, 온톨로지 서브 시스템은 시맨틱 웹에서 온톨로지를 생성하고 유지·관리하기 위한 시스템 구조이다(1). 그림 10은 OntoWeb 프로젝트에서 온톨로지 기반 정보검색 시스템의 기본적인 구조(30,36)를 설정한 것이다. 이들에게 알 수 있는 것은, 정보검색 시스템에서 온톨로지가 더 많은 관련 결과들을 검색하기 위한 가이드 역할을 담당한다는 것인데, 더 나아가 더 많은 관련 결과는 개념에 대한 의미를 컴퓨터가 정확히 이해한다는 전제가 따른다는 것을 알 수 있다. 이는 이전의 정보검색 시스템에서 해결하지 못한 질의나 문맥에서의 의미에 대한 처리 방법의 해결이라 할 수 있다.

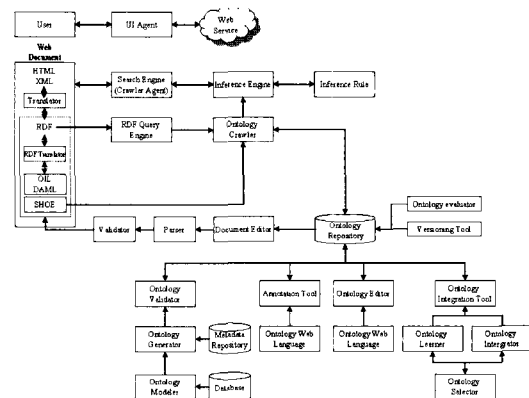


그림 9 Architecture of Semantic Web based Retrieval System

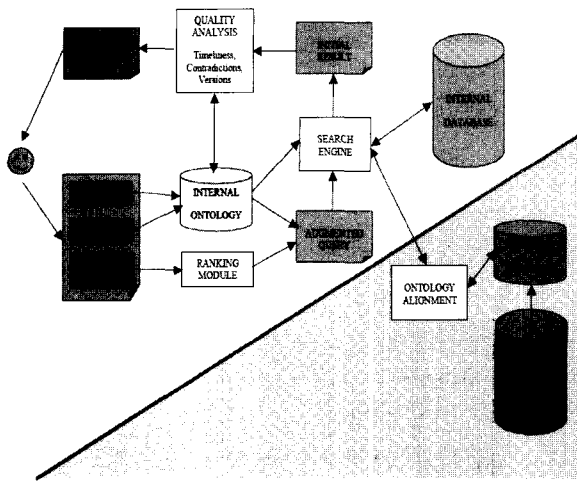


그림 10 Basic architecture for ontology based information retrieval applications

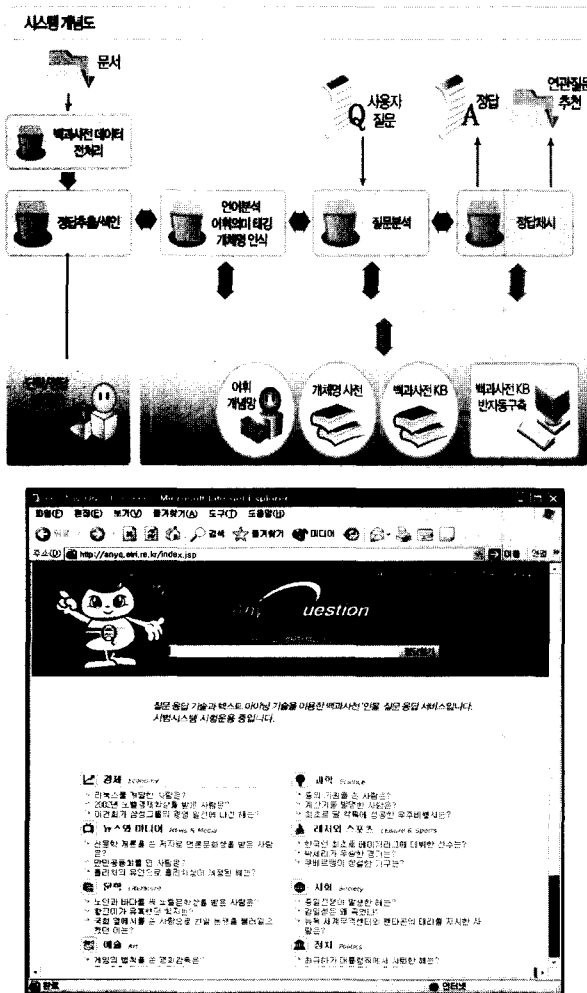


그림 11 "AnyQ" Question Answering System

ETRI에서 개발 중인 질의응답 시스템인 AnyQ (Any Question)는 엄밀한 의미에서는 시맨틱 웹 기술을 이용한 시스템은 아니지만, 백과사전을 기반으로 지식베이스를 구축하고, 이 지식베이스와 ETRI 어휘 개념망(온

톨로지)을 유기적으로 연결시켜 일반적인 온톨로지 활용 기법을 어느 정도 이용하고 있는 시스템이라 할 수 있다[5]. 이것은 '개념망+정답 문서 집합(answer set) + 속성 집합'으로 이루어진 지식베이스를 활용한 의미기반 정보검색 시스템에서도 반영되고 있다[3]. 즉 자연언어 처리 기술, 질의응답 처리 기술, 텍스트 마이닝 기술 등을 시스템의 기반 기술로 하여, 사용자의 질의 의도를 분석하고, 온톨로지와 더불어 사실 정보(facts)와 지식 정보로 구성된 지식베이스가 유기적으로 연결되어, 온톨로지의 기본 활용 방법을 사용함과 동시에, 개체명 인식 기술과 의미 분석 기술 등을 활용하여 정확한 답을 제시할 수 있는 온톨로지와 지식베이스 기반 질의응답 시스템이라 할 수 있다. 현재 백과사전 '인물' 범주를 중심으로 시범적으로 서비스 하고 있다(<http://anyq.etri.re.kr>).

OntoBroker는 웹 문서를 분석하기 위하여 온톨로지를 이용함과 동시에 온톨로지 기반 질의 처리를 하고 있는 시스템으로서[37], 확장된 HTML 문법을 사용자들이 문서를 온톨로지 구조로 마크업하기 위하여 제안되었으며, 사용자들이 OntoBroker Interface를 통해 수집된 지식 정보를 검색하고 문헌의 내용을 이해하기 위하여 그 지식 정보를 이용하게 한다. 여기에서 사용된 온톨로지는 정보 제공자와 탐색자를 위한 공통된 언어이며, 개념과 의미관계, 특정한 규칙으로 구성되어 있다. 그림 13은 OntoBroker의 구조를 나타낸 것이다.

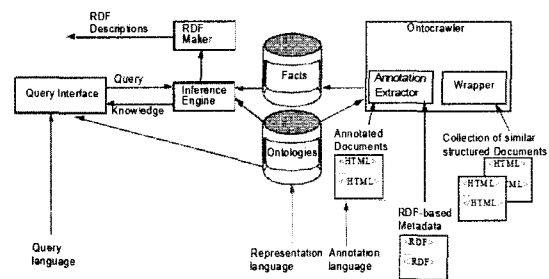


그림 12 OntoBroker Architecture

MELISA(Medical Literature Search Agent)는 의학 분야 문헌 검색 에이전트 시스템[26]으로서, 의학에서의 온톨로지 기반 정보검색 에이전트의 프로토타입이라 할 수 있다. 이 시스템은 질의 생성에 사용되는 의학 지식을 중심으로 구축된 Medical 온톨로지, 18,000 카테고리로 만들어진 MeSH(Medical Subject Headings) 온톨로지와 Terms 등을 사용하여 기존 의학 분야 문헌 검색 시스템의 성능을 향상시키고자 개발된 문헌 검색 시스템이라 할 수 있다. 사용자 질의에 대한 처리 문제와 의학 분야 문헌의 질의에 의한 문헌 검색 정확률 향상을 위해, 검색 엔진의 중요한 요소이자 질의 생성 모델에의 활용으로 Medical 온톨로지를 사용하고 있다.

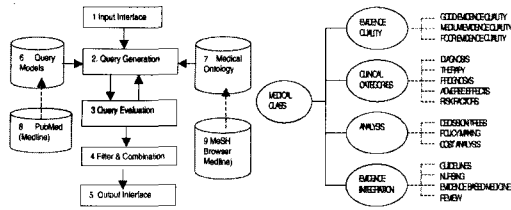


그림13 General structure of MELISA(left), and Classes & instances of the medical ontology(right)

OSE(Odyssey Search Engine) 시스템은 온톨로지를 사용한 특정 영역 지식(domain knowledge)에 기반한 검색 시스템[23]으로서, 분산되고 이질적인 지식 정보에 상관없이 다양한 영역에서 특정 영역의 정보를 제어할 수 있도록 하는 검색 에이전트 시스템 개발을 목적으로 하고 있다. 또한 domain 온톨로지는 사용자의 특정 영역 정보의 식별과 검색을 위한 시스템의 기본 개념들이다.

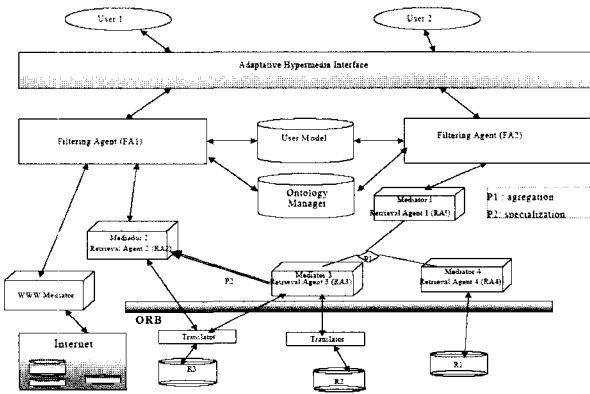


그림 14 Basic Architecture of Odyssey-Search Engine

이외에 온톨로지 기반 정보검색 시스템을 다음과 같이 정리할 수 있다.

온톨로지 기반 이미지 검색 시스템과 관련하여, 먼저 Helsinki University 박물관의 이미지 데이터베이스를 바탕으로 한 시맨틱 웹 기술을 이용한 검색 시스템 연구 개발[25]이 진행 중으로, 이 시스템은 이미지의 의미 주석과 검색을 위해 promotion 온톨로지에 기반한 이미지 검색 시스템 개발을 목표로 하고 있다. 이와 더불어 고대 가구 검색시스템에 사용하기 위하여 미술·건축 시소러스(AAT: Art and Architecture Thesaurus)의 고대 가구를 중심으로 RDFS를 이용한 온톨로지 구축하는 방법과 ATT와의 연결 방법을 제시한 연구[18]도 있다.

또한, 온톨로지 기반의 교차언어 정보검색시스템 개발도 최근 연구가 활발한데, 대표적인 연구 기관으로 New Mexico State University의 CRL(Computing Research Laboratory)이 유명하다. 이 연구실은 Mikro-

kosmos Ontology를 이용한 다국어 기계번역으로 유명한데, 온톨로지 기반의 교차 언어 정보검색(CLIR: Cross Language Information Retrieval) 시스템 연구 개발[20]이 진행 중에 있다. 국내에서도 온톨로지 기반 한의학 처방 관리시스템[6]이나 온톨로지 기반 웹 검색 시스템[7] 등이 개발되고 있으나 아직까지 실험적인 수준에 그치고 있다.

현재 이러한 일련의 온톨로지 기반 정보검색 시스템에 대한 평가 방법[24,31]에 대한 연구도 많이 진행되고 있다. 하지만 아직까지 정확한 평가 기준을 마련하기에는 어려운 실정이다.

4. 결 론

시맨틱 웹에서의 온톨로지 활용 기술은 웹 서비스 관련 분야 뿐만 아니라 생물정보학, 자연언어처리, 데이터베이스, 인공지능, 정보검색, 기계번역, 분산 시스템 등 다양한 분야에서 연구 대상으로 삼고 있다. 이것은 기존의 시소러스나 의미망, 어휘 데이터베이스 등과 같은 어휘의 계층적 구조와 어휘들의 관계로 표현되었던 어휘 집합 체계를 특정 분야에서만 활용했던 것과는 사뭇 다르다. 즉 온톨로지가 갑자기 관련 분야의 핵심 연구 대상으로서 차세대 정보 처리의 핵심 기술이나 자원으로 부각된 이유는, 먼저 웹이라는 거대한 지식 정보의 처리 방법과 현재 우리가 처한 수많은 지식 정보의 체계화 방안이 어느 정도 일맥상통함과 동시에, 다음으로 그 지식 정보의 처리에 사람의 사고방식과 동일한 형태의 의미 정보를 컴퓨터에 담아, 컴퓨터가 의미를 이해할 수 있도록 만든다는 메커니즘 때문이라고 판단된다.

현재 온톨로지에 대한 연구는 국내외적으로 지식기반·지능형 시스템 개발에서는 제외할 수 없는 핵심 기술이자 자원으로 인식되고 있다. 또한 많은 연구자들이 이 온톨로지에 관심을 가지고 있는 것도 사실이다. 하지만 국내에서는 아직까지 국외의 온톨로지 관련 프로젝트와 연구 활동처럼 근본적인 기반 연구가 되어 있지 못한 실정이다. 정보 검색 시스템 뿐만 아니라 다른 온톨로지 활용 시스템 개발을 위해서는 이러한 기반 연구가 선행되어야 하며, 국가적 차원의 연구 기관과 대규모 프로젝트를 통해 국내외 온톨로지 표준화 작업의 참여뿐만 아니라, 시맨틱 웹 기술의 국내 기반 확충과 성장, 국제 표준에의 참여를 위해 많은 연구자들이 노력해야 할 것이다.

참고문헌

[1] 이재호, "시맨틱 웹의 온톨로지 언어", 정보과학회지, 제21권, 제3호, pp. 18-27, 2003

- [2] 양정진, "시맨틱 웹에서의 온톨로지 공학", 정보과학회지, 제21권, 제3호, pp.28-35, 2003
- [3] 장명길 외, "의미기반 정보검색", 정보과학회지, 제19권, 제10호, pp. 7-18, 2001
- [4] 옥철영, "한국어정보처리와 온톨로지", 2004 한국어정보처리연구회 동계 튜토리얼 자료집
- [5] 최호섭, 옥철영, 김창환, 왕지현, 장명길, "질의응답 시스템을 위한 백과사전 기반 지식베이스와 온톨로지", 제15회 한글 및 한국어 정보처리학술대회 자료집, pp. 177-183, 2003
- [6] 이현실, 이두영, "온톨로지 기반 한의학 처방 지식 관리시스템 설계에 관한 연구", 정보관리학회지, 제20권, 제1호, pp. 341-371, 2003
- [7] 김현희, 안태경, "온톨로지를 이용한 인터넷 웹 검색에 관한 실험적 연구", 정보관리학회지, 제20권, 제1호, pp. 417-455, 2003
- [8] 정도현, "시맨틱 웹을 위한 온톨로지 언어와 구현 사례 연구", 정보관리연구, 제34권, 제3호, pp. 87-109, 2003
- [9] 이강찬, 김성한, 민재홍, 박기식, 정인정, "시맨틱 웹 기반의 검색 시스템 구조", 주간기술동향, 제1094호, IITA IT 정보단, 2003, <http://www.itfinder.or.kr>
- [10] Berners-Lee, T., Hendler, J., Lassila, O., "The Semantic Web," Scientific American, 2001
- [11] Maedche A., Ontology Learning for the Semantic Web, Academic Publishers, 2002
- [12] Davies, J., Fensel, D., Harmelen, F.V., Towards The Semantic Web, JOHN WILEY & SON Ltd, 2003
- [13] Fensel, D., Hendler, J., Lieberman, H., Wahlster, W., Spinning the Semantic Web, The MIT Press, 2003
- [14] Latifur Kahn, Feng Luo, "Ontology Construction for Information Selection," Proceedings of the 14th IEEE International Conference on Tools with Artificial Intelligence, pp. 122-127, 2002
- [15] Daconta, M.C., Obrst, L.J., Smith, K.T., The Semantic Web, Wiley Publishing Inc. 2003
- [16] Paola Velardi, Paolo Fabriani, Michele Missikiouf, "Using Text Processing Techniques to Automatically Enrich a Domain Ontology," Proceedings of the international conference on Formal Ontology in Information System, pp. 270-284, 2001
- [17] Gloria L. Zúñiga, "Ontology: its transformation from philosophy to information systems," Proceedings of the international conference on Formal Ontology in Information System, pp. 187-197, 2001
- [18] Wielinga, B.J., Schreiber, A.Th., Wielemaker, J., Scandberg, J.A.C., "From Thesaurus to Ontology," Proceedings of the international conference on Knowledge capture, pp. 194-201, 2001
- [19] Yannis Tzitzikas, Collaborative Ontology-based Information Indexing and Retrieval, Doctoral Dissertation, Department of computer science, University of Crete, 2002
- [20] Abdelali. A., Cowie, J., Farwell, D., Ogden, B., Helmreich, S., "Cross-Language Information Retrieval using Ontology," TALN 2003, 2003
- [21] José Saias, Paulo Quaresma, "A methodology to create ontology-based information retrieval systems," EPIA'03-11th Portuguese Conference on Artificial Intelligence, pp. 424-434, 2003
- [22] Guarino, N., Giarretta, Pierdaniele., "Ontology and Knowledge Bases-Toward a Terminological clarification," In N. Mars (ed.), Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing, pp. 25-32, IOS Press, 1995
- [23] Braga, R.M.M., Werner, C.M.L., Mattoso, M., "Using Ontologies for Domain Information Retrieval," 11th International Workshop on Database and Expert Systems Applications (DEXA'00), pp. 836-840, 2000
- [24] Aitken, S., Reid, S., "Evaluation of an Ontology-Based Information Retrieval Tool," 12th European conference on Artificial Intelligence(ECAI'00) Workshop on Applications of Ontologies and Problem-Solving Method, 2000
- [25] Hyvönen E., Styrman A., Saarela, S., "Ontology-based Image Retrieval," Number

2002-03 in HIIT Publications, pp. 15-27, Helsinki Institute for Information Technology(HIIT), 2002

[26] Abasolo, J.M., Gómez, M., "MELISA. An Ontology-based agent for information retrieval in medicine," ECDL 2000 Workshop on the Semantic Web(SemWeb2000), pp. 73-82, 2000

[27] Gruber, T., "A Translation approach to portable ontology specifications," Knowledge Acquisition, vol.5, no.2, pp. 199-220, 1993

[28] Gruber, T., "Toward Principles for the design of ontologies used for knowledge sharing," International Journal of Human-Computer Studies, vol.43, no.5/6, pp. 907-928, 1995

[29] Uschold, M., Gruning, M., "ONTOLOGIES: Principles, Methods, and Applications," AIAI-TR-191, Artificial Intelligence Applications Institute(AIAI), the University of Edinburgh, 1996

[30] Uschold, M., King, M., Moralee, S., Zorgios, Y., "The Enterprise Ontology," AIAI-TR-195, Artificial Intelligence Applications Institute (AIAI), the University of Edinburgh, 1997

[31] IST Project IST-2000-29243 OntoWeb report, Project Full Title: Ontology-based information exchange for knowledge management and electronic commerce

[32] Moench, E., Ullrich, M., Schnurr, H., Angele, J., "SemanticMiner: Ontology-based Knowledge Retrieval," whitepaper, <http://www.ontolprise.de>

[33] <http://www.w3c.org>

[34] <http://www.daml.org>

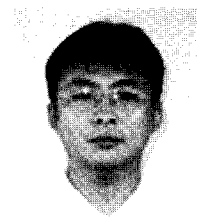
[35] <http://www.ontoknowledge.org>

[36] <http://www.cs.umd.edu/projects/plus/SHOE/>

[37] <http://ontoweb.aifb.uni-karlsruhe.de>

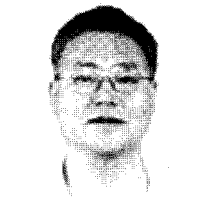
[38] <http://tap.stanford.edu>

최 호 섭



1998 경남대학교 국어국문학과 문학사
 2000 경남대학교 국어국문학과 문학석사
 2000~2001 한국전자통신연구원 지식정보검색연구팀 과건연구원
 2002~2003 (주)시리울산 한국어시소러스연구소 선임연구원
 2002~현재 울산대학교 컴퓨터정보통신공학과 석박사통합과정
 관심분야 : 한국어정보처리, 온톨로지, 지식베이스, 언어이해, 지식처리

옥 철 영



1982 서울대학교 컴퓨터공학과 학사
 1984 서울대학교 컴퓨터공학과 석사
 1993 서울대학교 컴퓨터공학과 박사
 1994 러시아 TOMSK 공과대학 교환교수
 1996 영국 GLASCOW 대학교 객원교수
 1984~현재 울산대학교 컴퓨터정보통신공학부 교수
 관심분야 : 한국어정보처리, 의미분별, 온톨로지, 지식베이스, 기계학습, 문서분류
 E-mail : okcy@mail.ulsan.ac.kr

• The 6th IEEE Int'l Conference on Automatic Face and Gesture Recognition(FG 2004) •

- 일 자 : 2004년 5월 17~19일
- 장 소 : 롯데호텔(서울)
- 주 최 : 컴퓨터비전및패턴인식연구회
- 상세안내 : <http://image.korea.ac.kr/FG2004>